

School of Information and Library Science  
University of North Carolina at Chapel Hill  
December 3, 2001

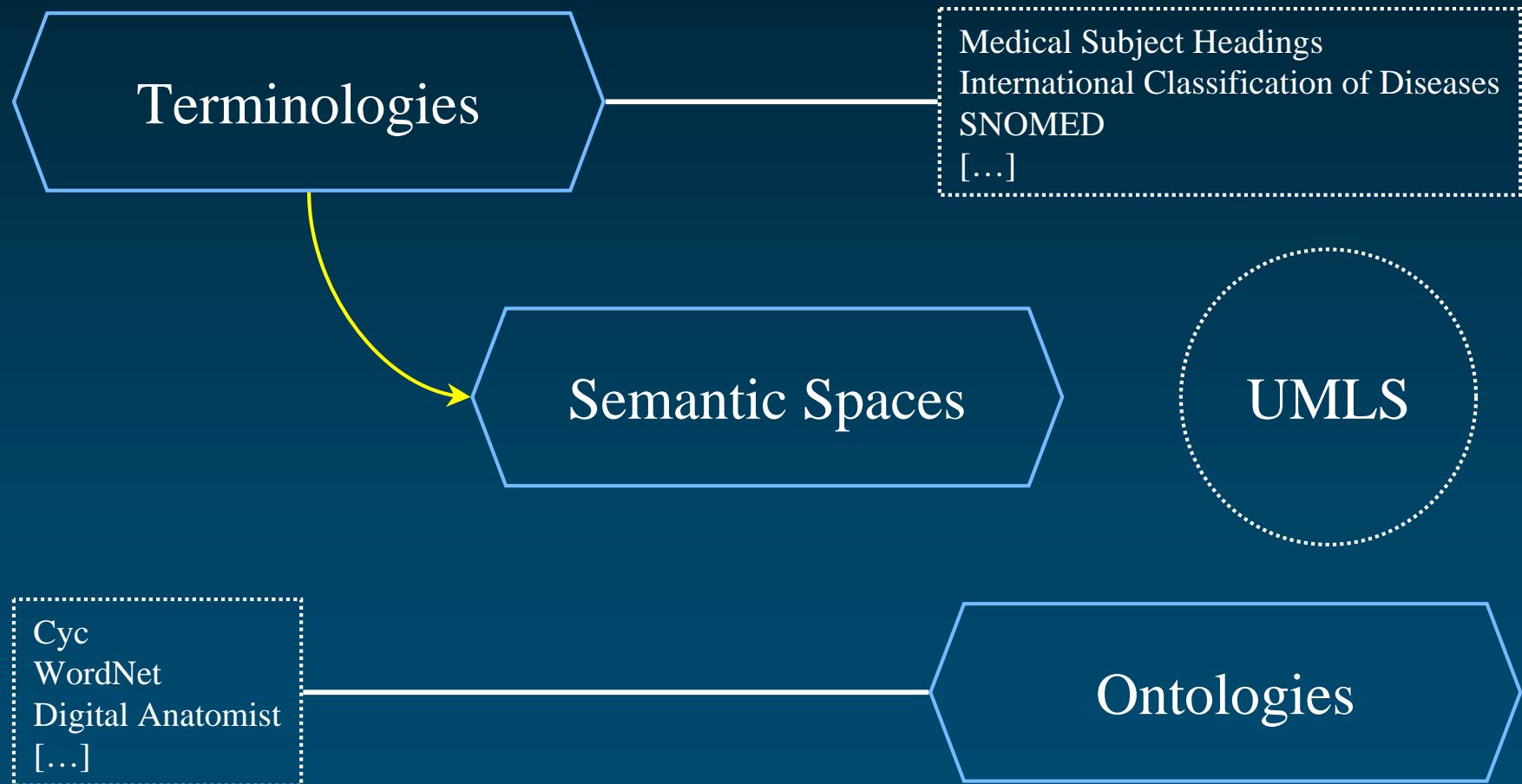
## Issues in the organization of biomedical knowledge



*Olivier Bodenreider, MD, PhD*

Lister Hill National Center  
for Biomedical Communications  
Bethesda, Maryland - USA

# Biomedical knowledge organization



# Biomedical terminologies

---

## ◆ Core vocabularies

- anatomy (UWDA, Neuronames)
- drugs (First DataBank, Micromedex)
- medical devices (UMD, SPN)

## ◆ Several perspectives

- clinical terms (SNOMED, CTV3)
- information sciences (MeSH, CRISP)
- administrative terminologies (ICD-9-CM, CPT-4)
- standards (HL7, LOINC)



# Biomedical terminologies (cont'd)

---

- ◆ Specialized vocabularies
  - nursing (NIC, NOC, NANDA, Omaha, PCDS)
  - dentistry (CDT)
  - oncology (PDQ)
  - psychiatry (DSM, APA)
  - adverse reactions (COSTART, WHO ART)
  - primary care (ICPC)
- ◆ Knowledge bases (AI/Rheum, DXplain, QMR)

# UMLS

---

Unified Medical Language System

Unified Medical Language System



# UMLS Metathesaurus

---

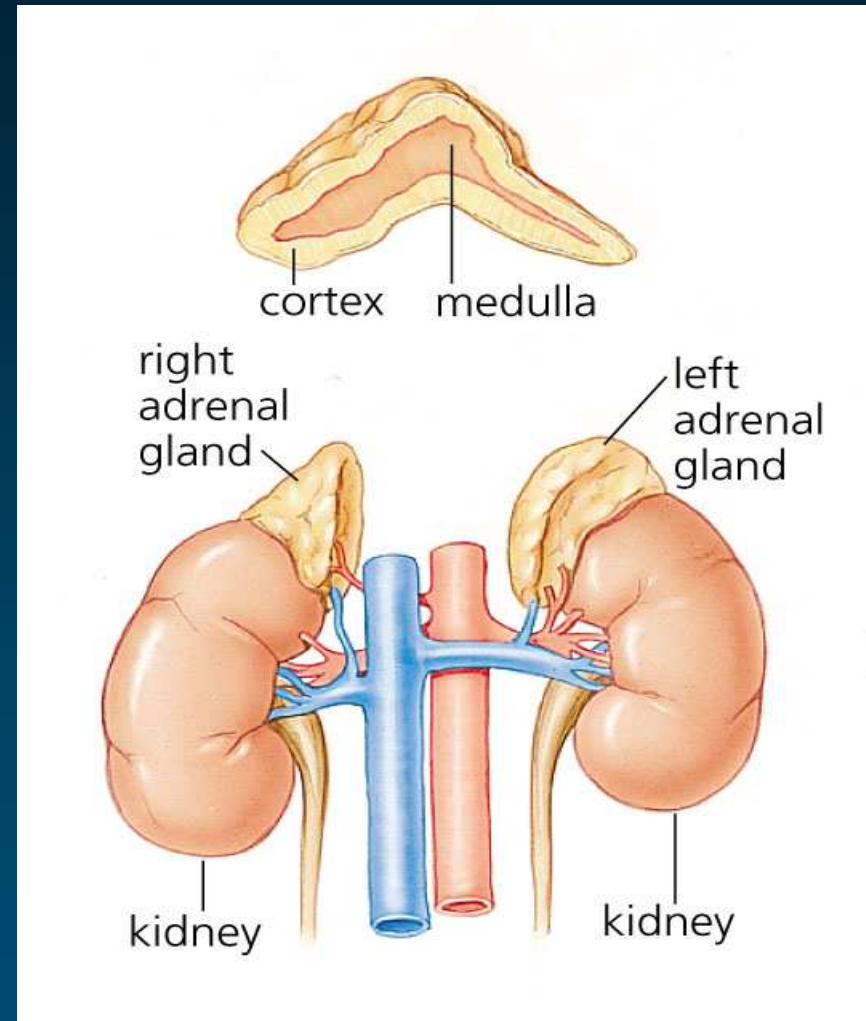
- ◆ 50 families of vocabularies
- ◆ 800,000 concepts
- ◆ 1.7 M names
- ◆ About 10 M inter-concept relationships
  - Symbolic
  - Statistical



# Naming and Organizing

# Addison's disease

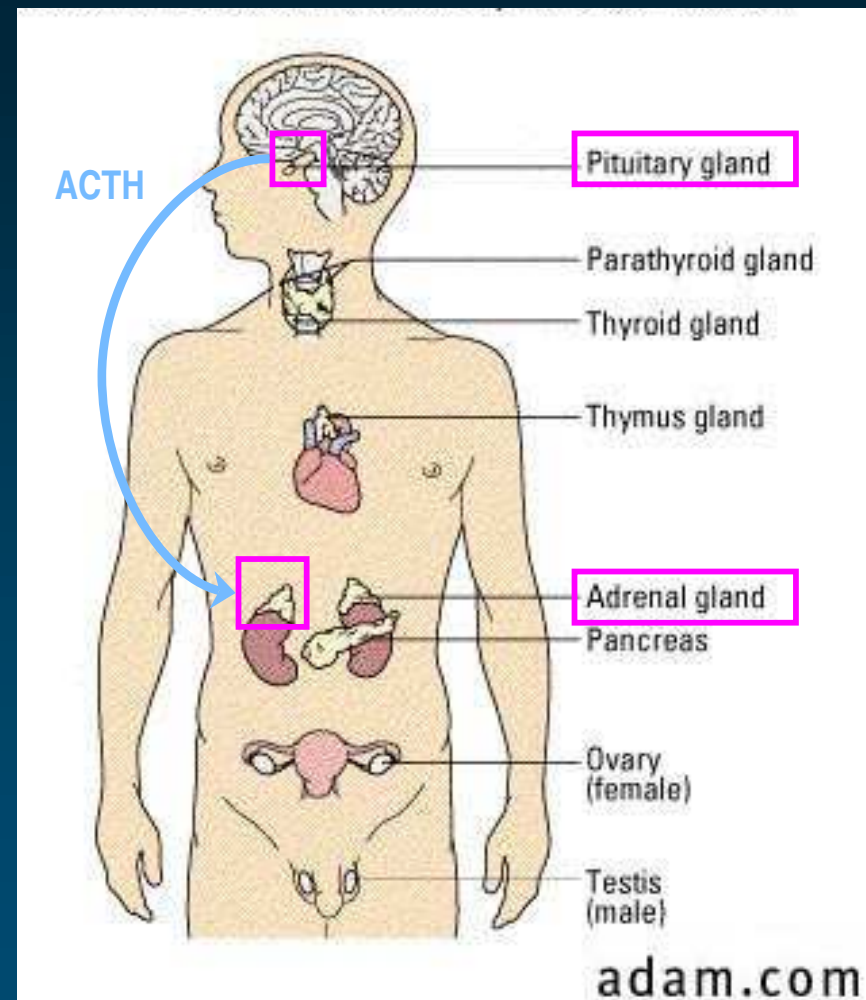
- ◆ Addison's disease is a rare endocrine disorder
- ◆ Addison's disease occurs when the adrenal glands do not produce enough of the hormone cortisol
- ◆ For this reason, the disease is sometimes called chronic adrenal insufficiency, or hypocortisolism





# Adrenal insufficiency Clinical variants

- ◆ Primary / Secondary
  - Primary: lesion of the adrenal glands themselves
  - Secondary: inadequate secretion of ACTH by the pituitary gland
- ◆ Acute / Chronic
- ◆ Isolated / Polyendocrine deficiency syndrome



# Addison's disease Symptoms

---

- ◆ Fatigue
- ◆ Weakness
- ◆ Low blood pressure
- ◆ Pigmentation of the skin (exposed and non-exposed parts of the body)
- ◆ ...

# AD in medical vocabularies

---

- ◆ Different names
- ◆ Different contexts



# AD in medical vocabularies Names

---

## ◆ Synonyms: different terms

- |  |   |                      |
|--|---|----------------------|
| ● Addisonian syndrome                  | } | eponym               |
| ● Bronzed disease                      |   |                      |
| ● Addison melanoderma                  | } | symptoms             |
| ● Asthenia pigmentosa                  |   |                      |
| ● Primary adrenal deficiency           | } | clinical<br>variants |
| ● Primary adrenal insufficiency        |   |                      |
| ● Primary adrenocortical insufficiency |   |                      |
| ● Chronic adrenocortical insufficiency |   |                      |

# AD in UMLS Names

- ◆ Synonymous terms clustered into a concept
- ◆ Preferred term
- ◆ Unique identifier (CUI)

Adrenal gland diseases	MeSH	D000307
Adrenal disorder	AOD	0000005418
Disorder of adrenal gland	Read	C15z.
Diseases of the adrenal glands	SNOMED	DB-70000

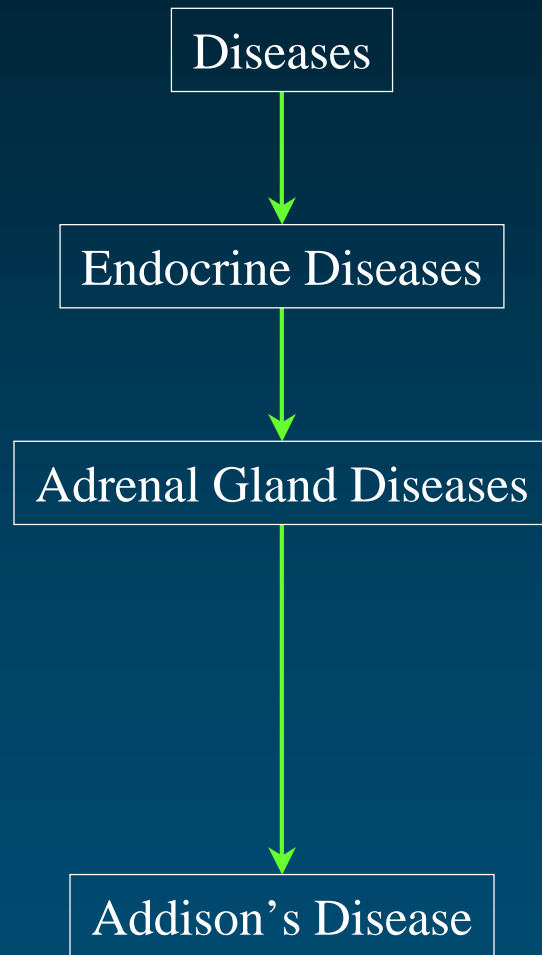
C0001621

Adrenal Gland Diseases

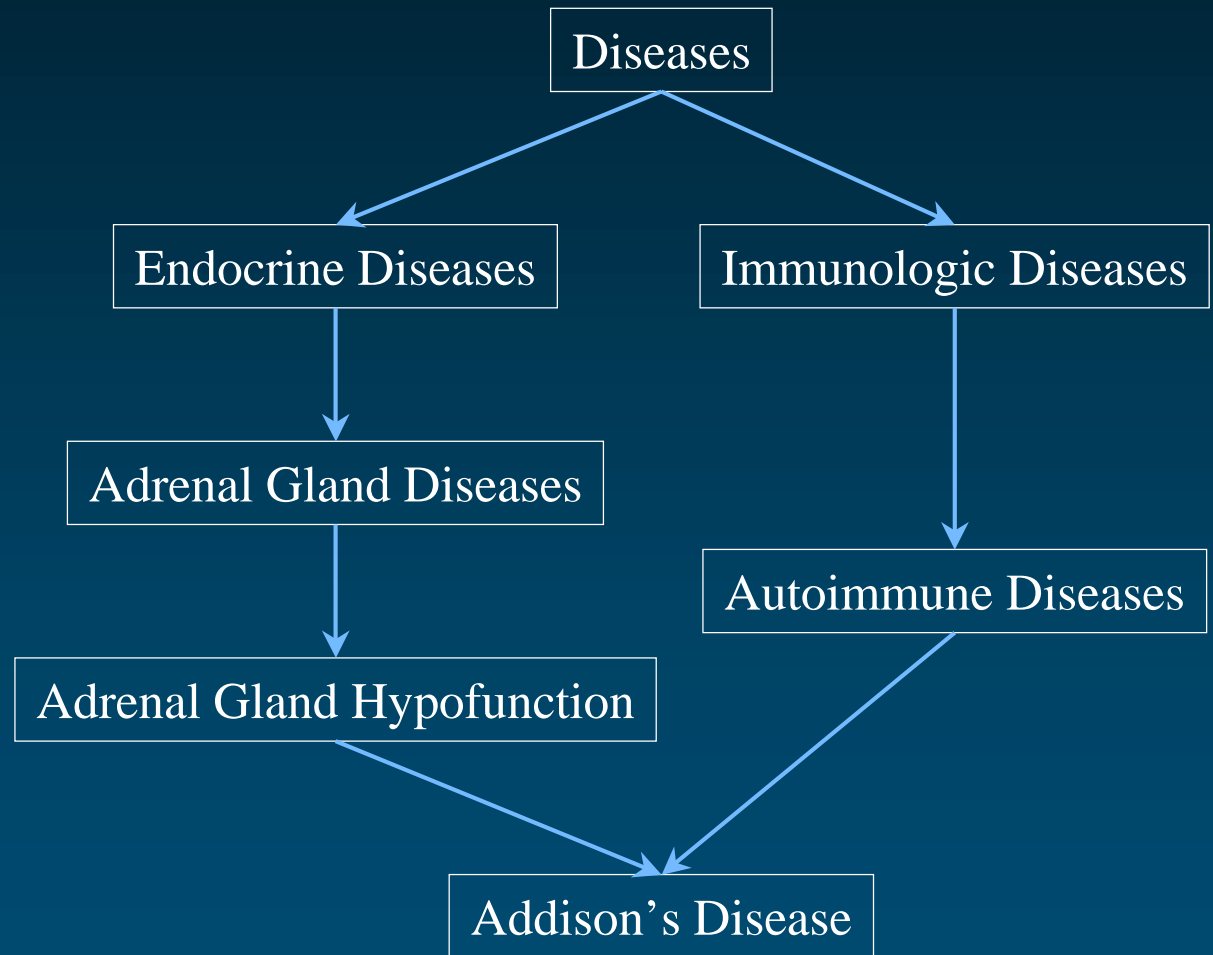


# AD in medical vocabularies Contexts

## SNOMED

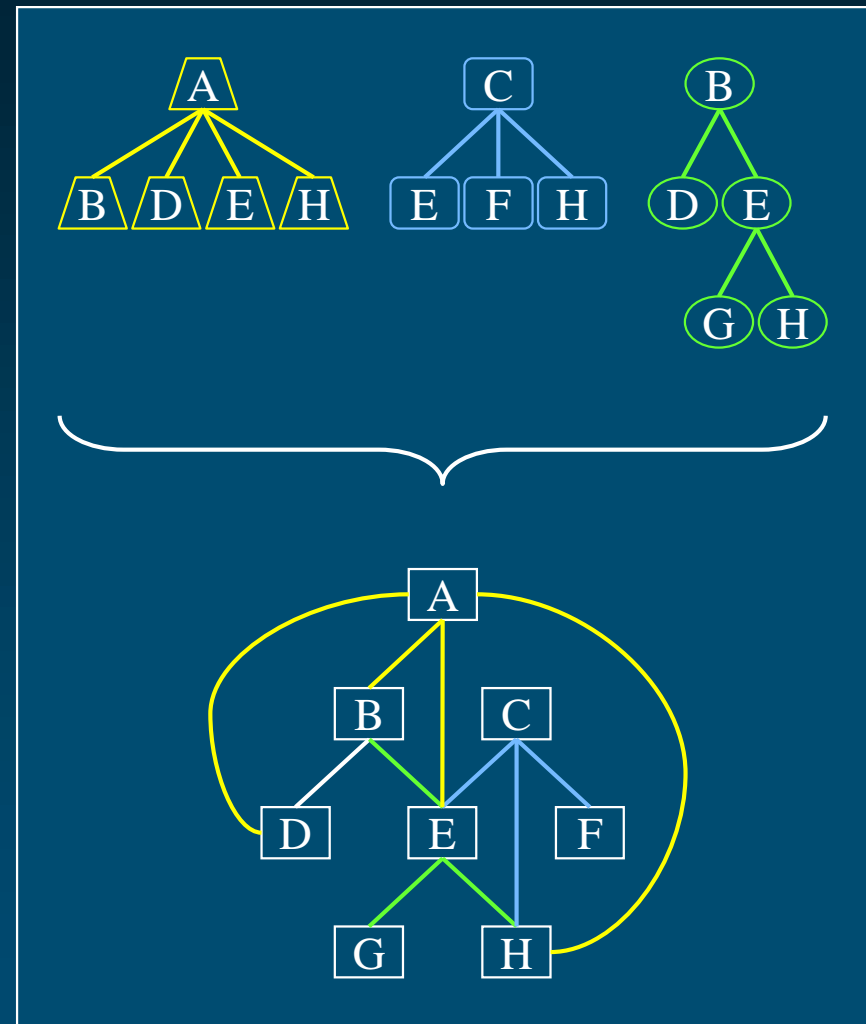


## MeSH

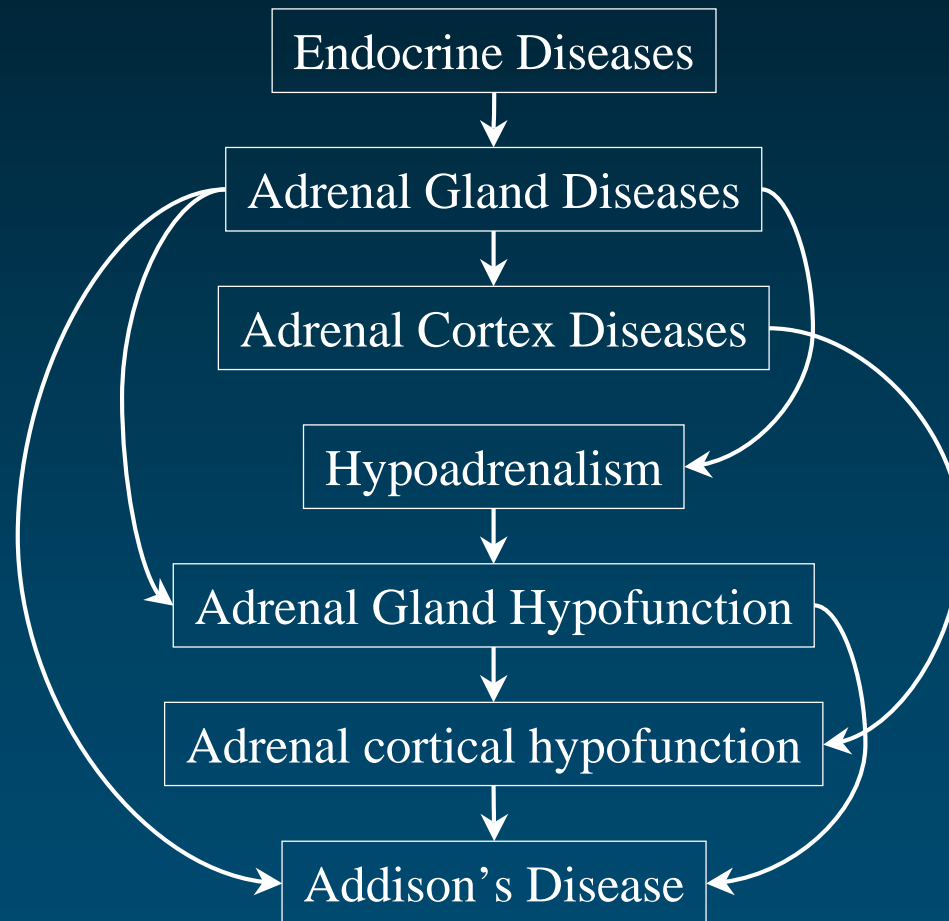


# AD in UMLS Contexts

- ◆ Multiple **tree** structures combined into a **graph** structure
- ◆ Directed **acyclic** graph (DAG)

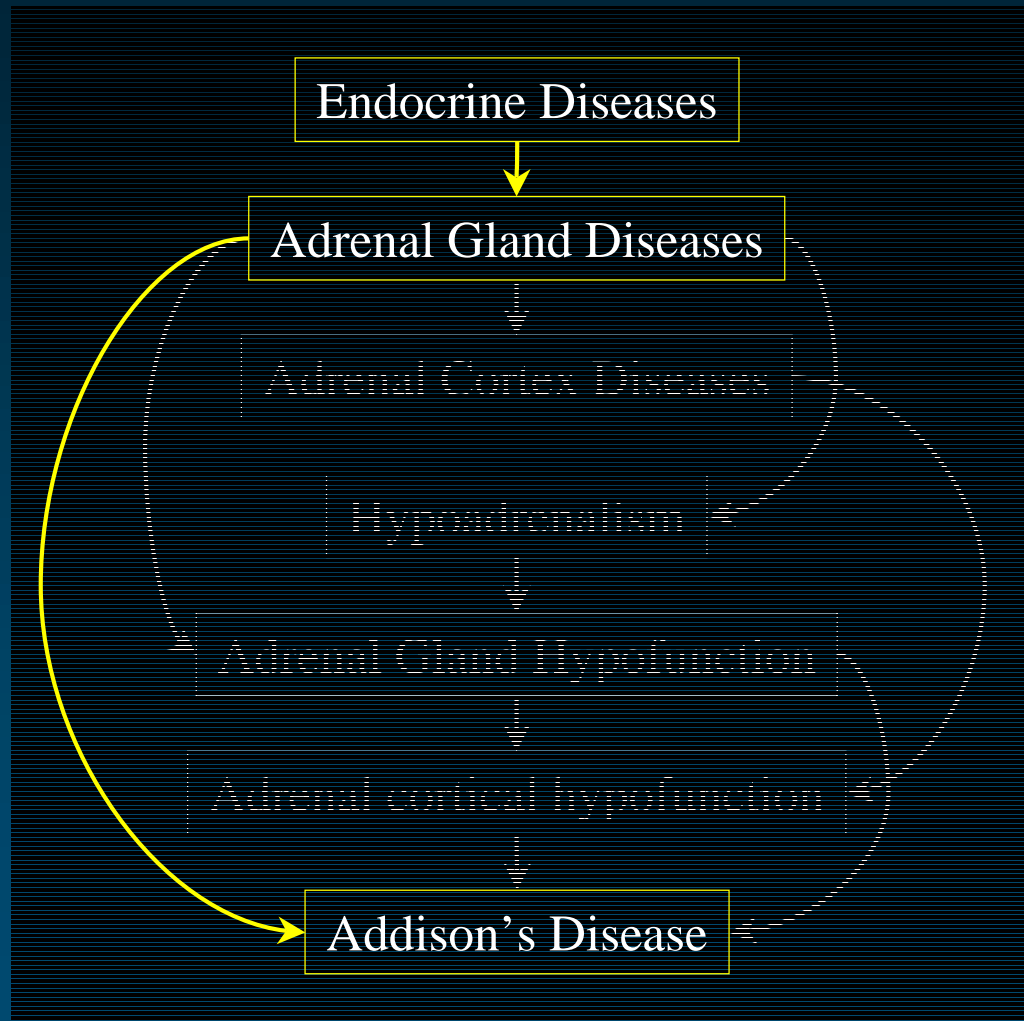


# AD in UMLS Contexts

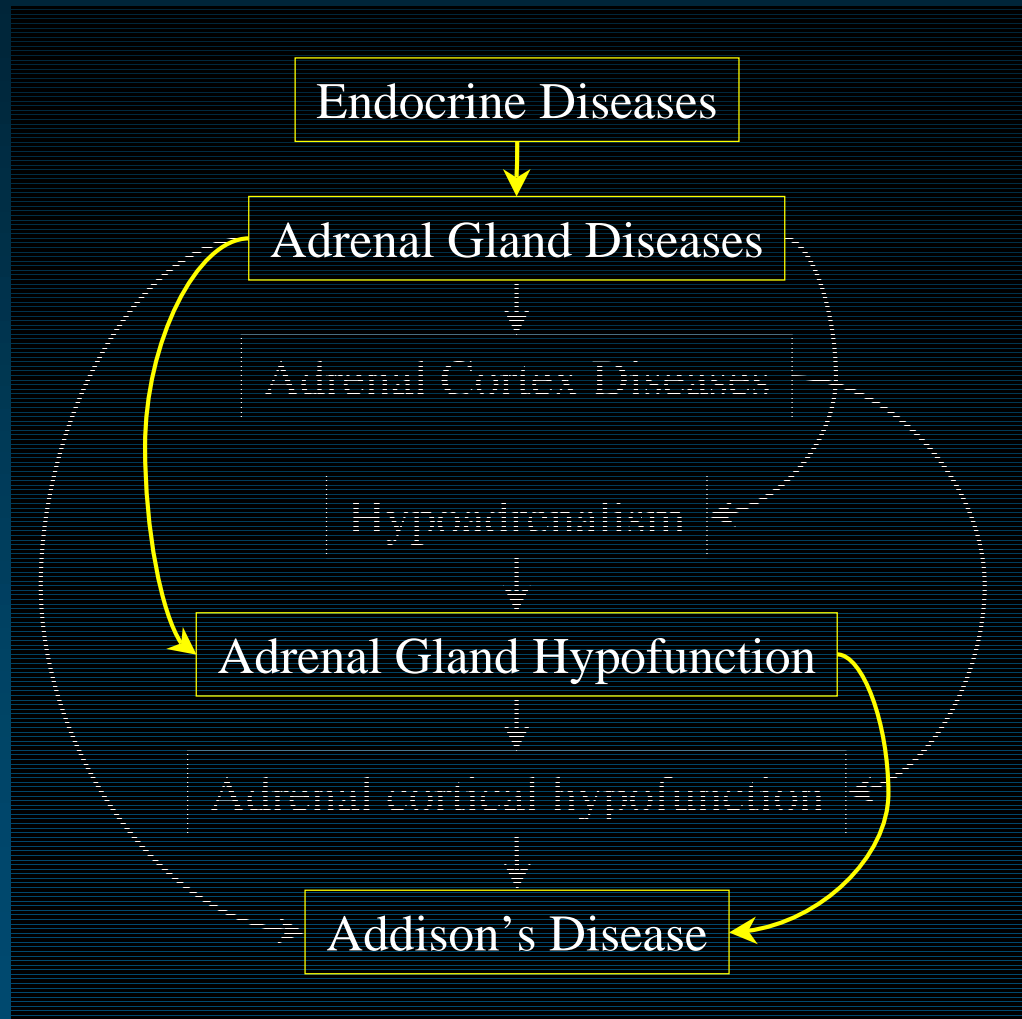




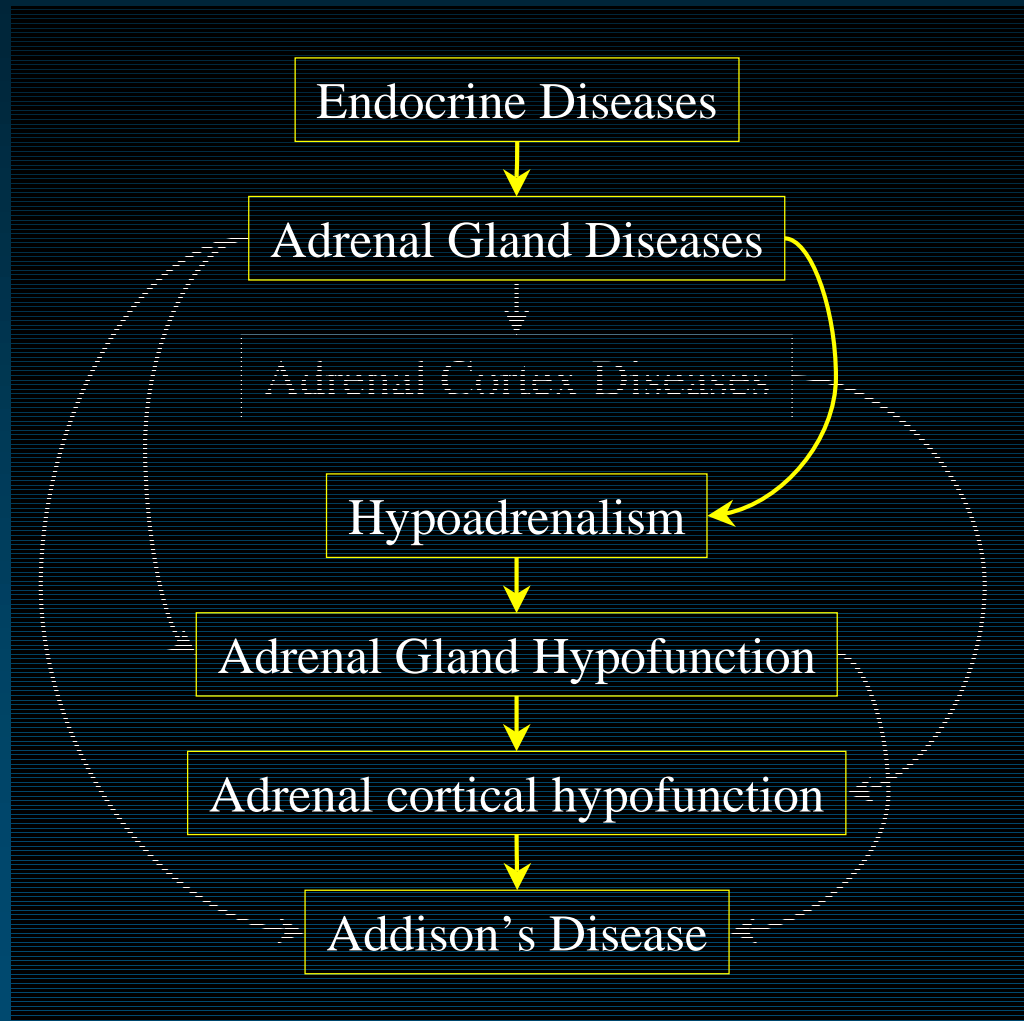
# AD in UMLS SNOMED context



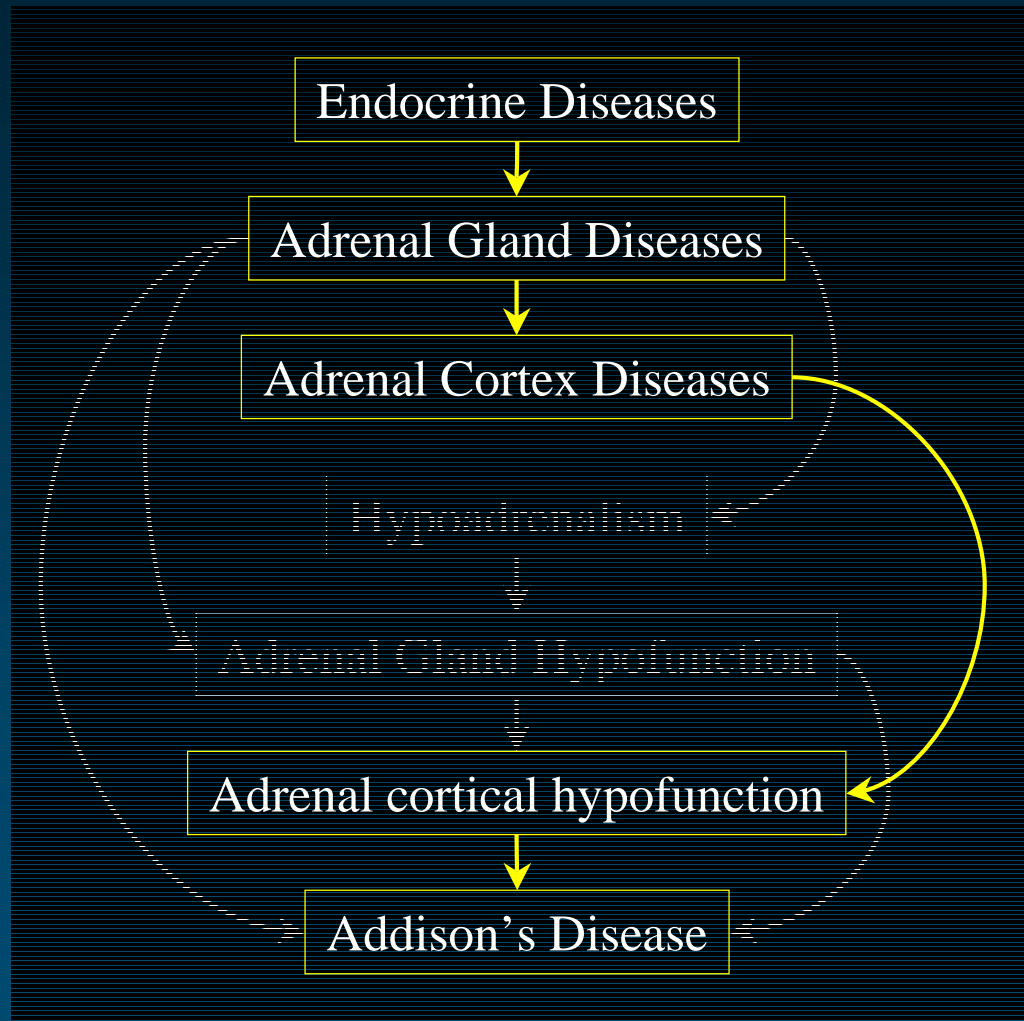
# AD in UMLS MeSH context

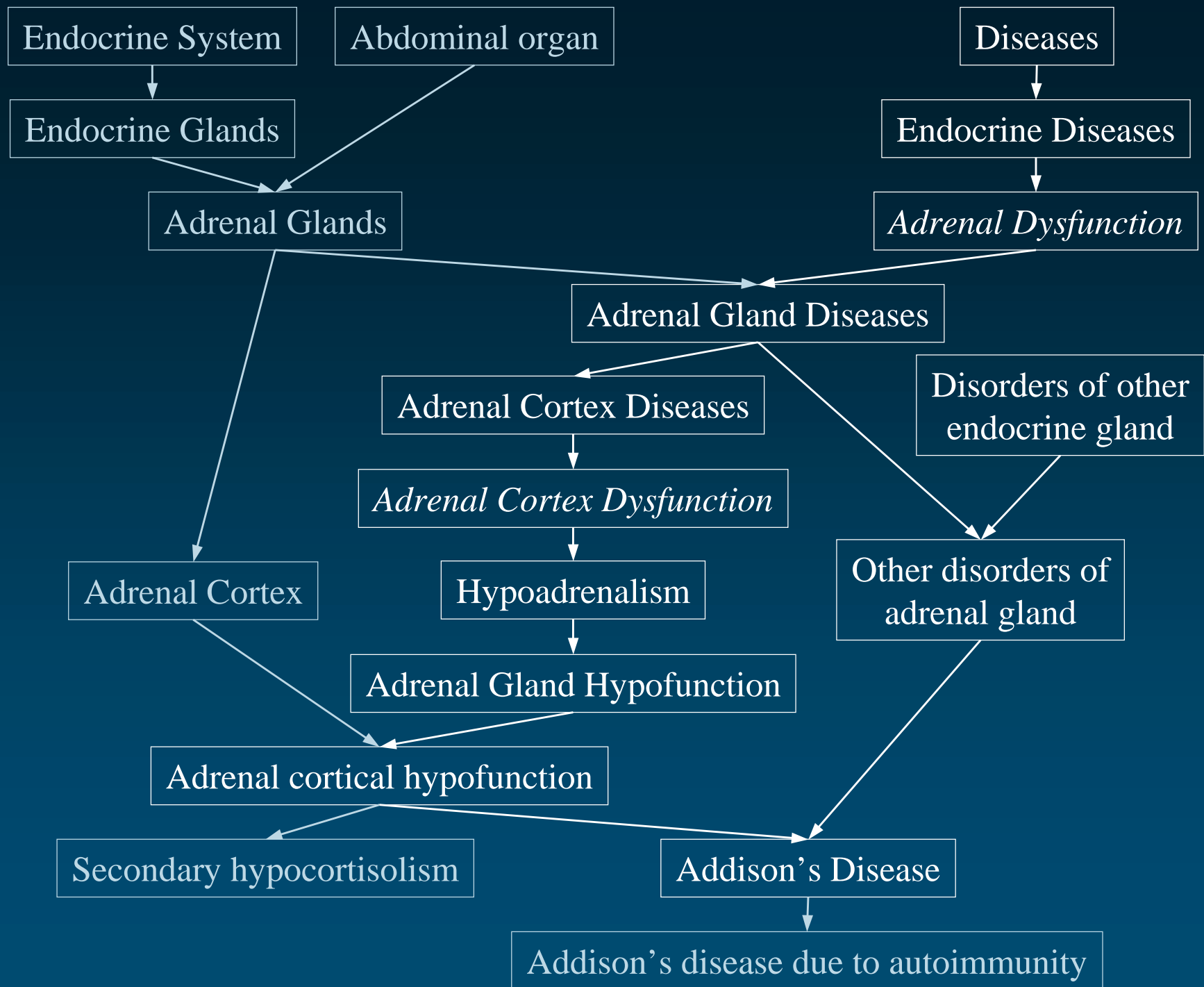


# AD in UMLS Read Codes context



# AD in UMLS AOD Thes. context





# Biomedical terms Complexity

---

## ◆ Mostly precoordinated

Autoimmune disorders

Addison's disease

Addison's disease due to autoimmunity

## ◆ Various complexity

Motor vehicle traffic accident involving collision, on highway, between motor vehicle and fallen stone, traffic sign, tree or utility pole, passenger of motor vehicle injured

## ◆ Neoclassical compounds

Open prostatectomy

# Biomedical terms Variability

---

- ◆ Case
- ◆ Punctuation
- ◆ Underspecification feature
- ◆ Genitive mark
- ◆ Inflection
- ◆ Word order

# Normalization Example

Hodgkin Disease  
HODGKINS DISEASE  
Hodgkin's Disease  
Disease, Hodgkin's  
Hodgkin's, disease  
HODGKIN'S DISEASE  
Hodgkin's disease  
Hodgkins Disease  
Hodgkin's disease NOS  
Hodgkin's disease, NOS  
Disease, Hodgkins  
Diseases, Hodgkins  
Hodgkins Diseases  
Hodgkins disease  
hodgkin's disease  
Disease, Hodgkin

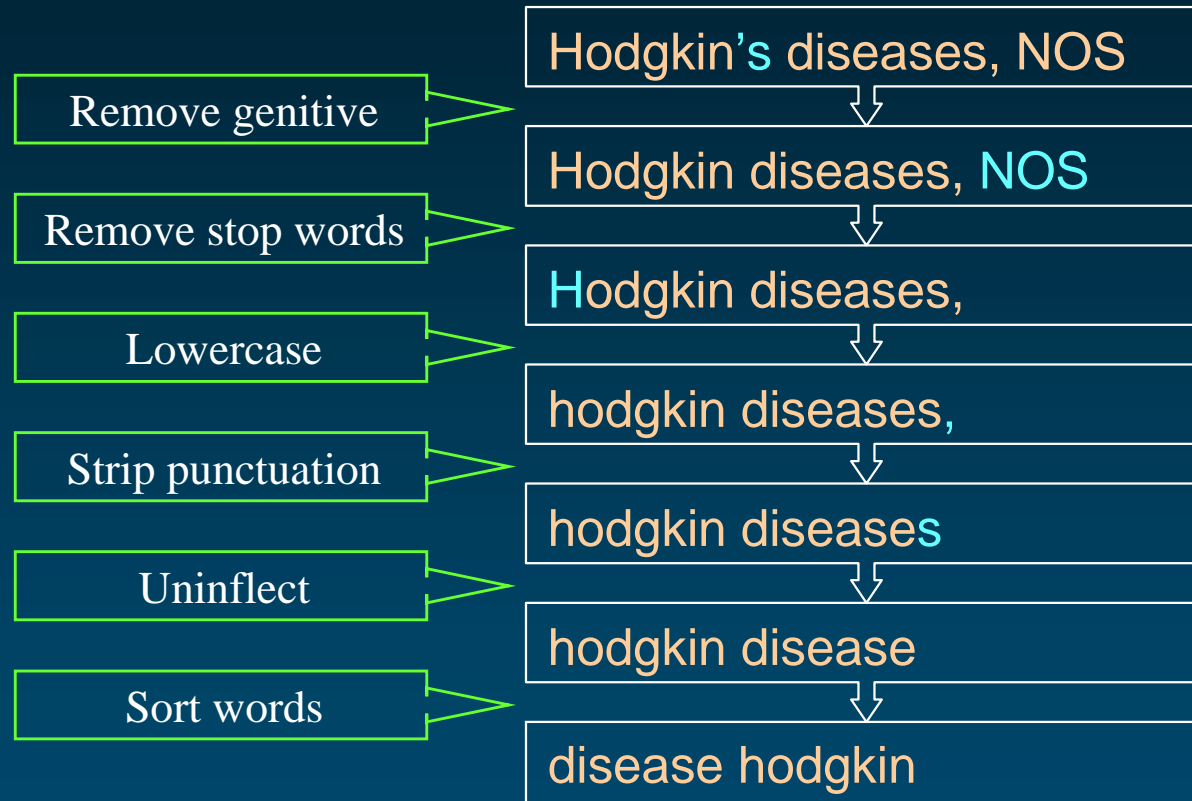
normalize

disease hodgkin

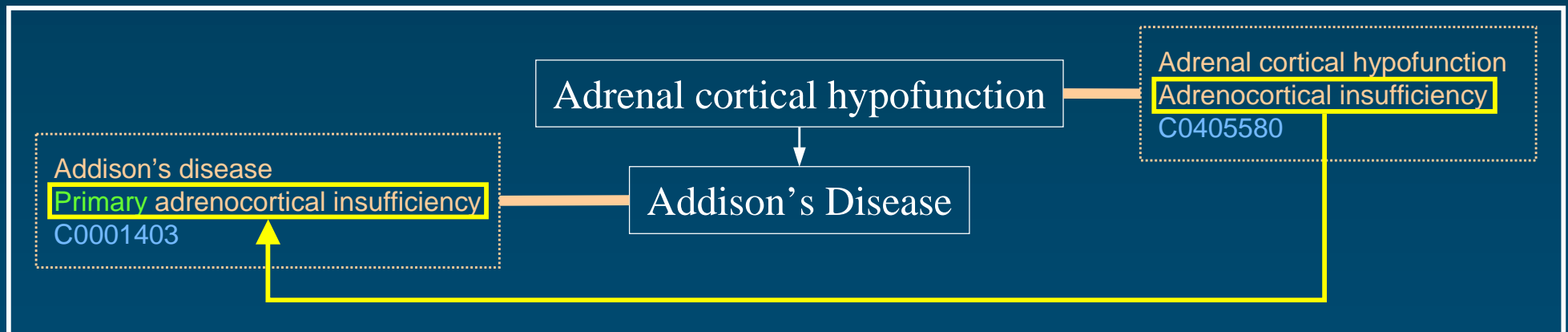
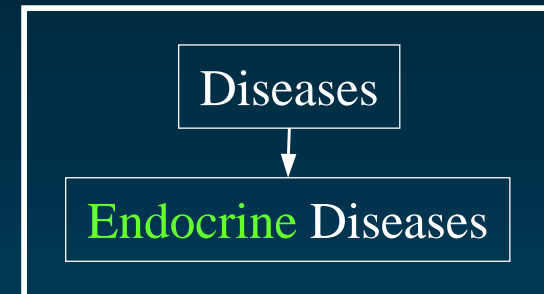
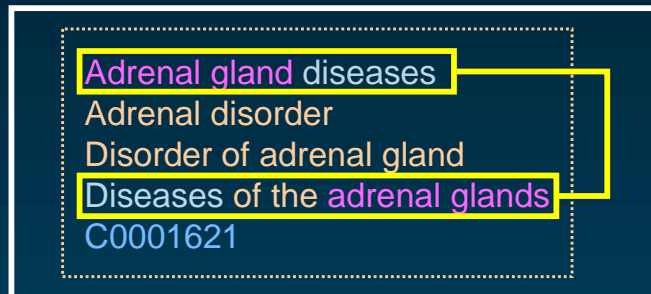




# Normalization



# Lexical knowledge



# Lexical resources

## ◆ Lexicon

Syntactic Category: noun  
Inflection Type: reg

Base Form: gland

Singular: gland  
Plural: glands

## ◆ Lexical tools

- stop words

Diseases of the adrenal glands

- word order

Diseases of the adrenal glands  
Adrenal glands diseases

- inflection

gland → glands

- derivation

cortex → cortical



# Synonymy

# From terms to concepts

---

- ◆ Terms are organized by meaning
- ◆ Clusters of synonymous terms
  - UMLS (concepts)
  - WordNet (synsets)
- ◆ Concepts are defined by extension
- ◆ Polysemous terms may belong to several concepts

# Synonymy and Granularity

---

## ◆ Splitters vs. Lumpers

- Formal vs. practical definition of synonymy

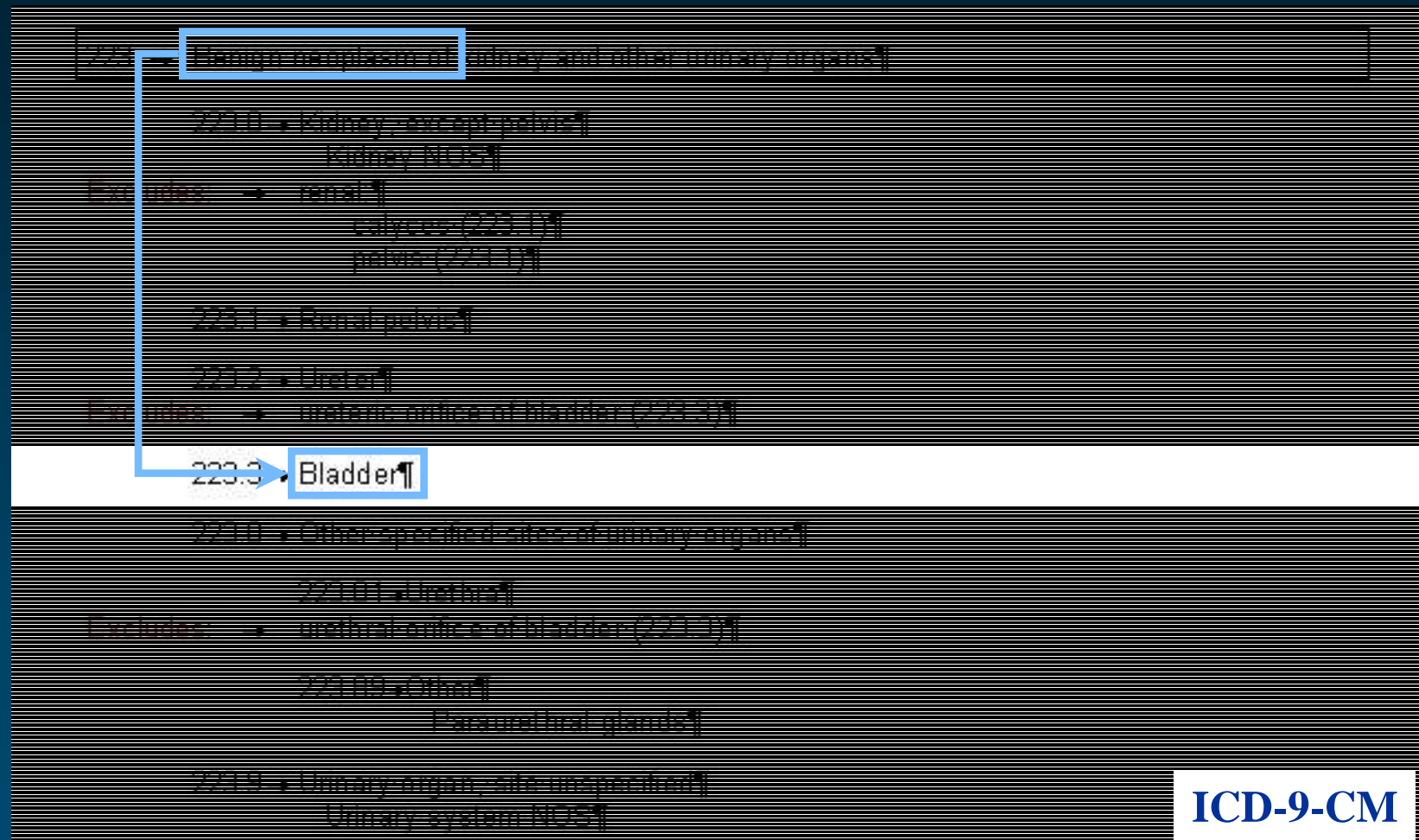
Posttransfusion hepatitis  
Posttransfusion **viral** hepatitis

## ◆ Organization vs. Representation

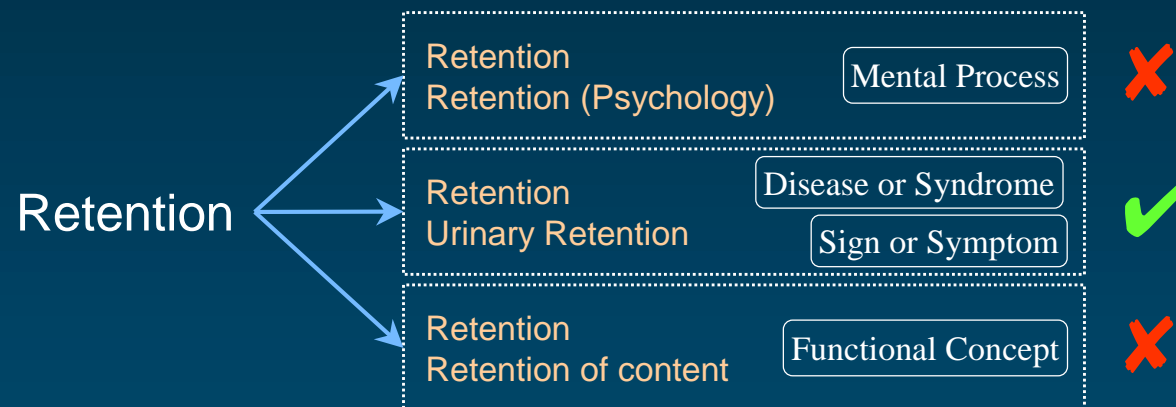
- Task-oriented vs. Purpose-independent
- Generic vs. prototypical

Asplenia  
**Congenital** asplenia

# Implicit contextual knowledge



# Disambiguate according to the context





# Hierarchical Relationships

# Hierarchy

## ◆ Hierarchical relationships

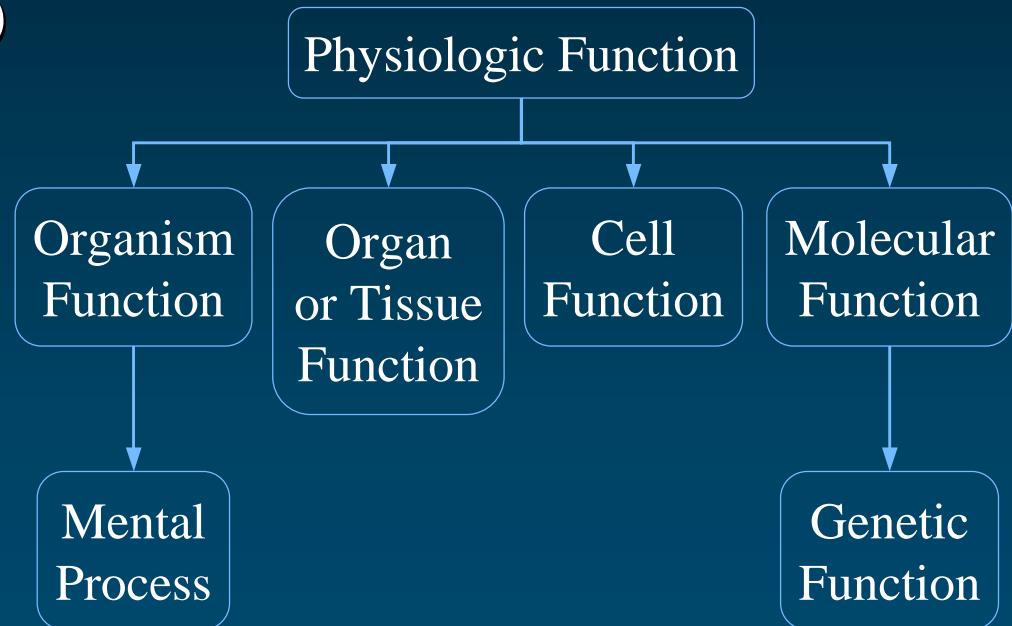
- Taxonomy (isa)
- Meronymy (part of)

## ◆ Partial ordering

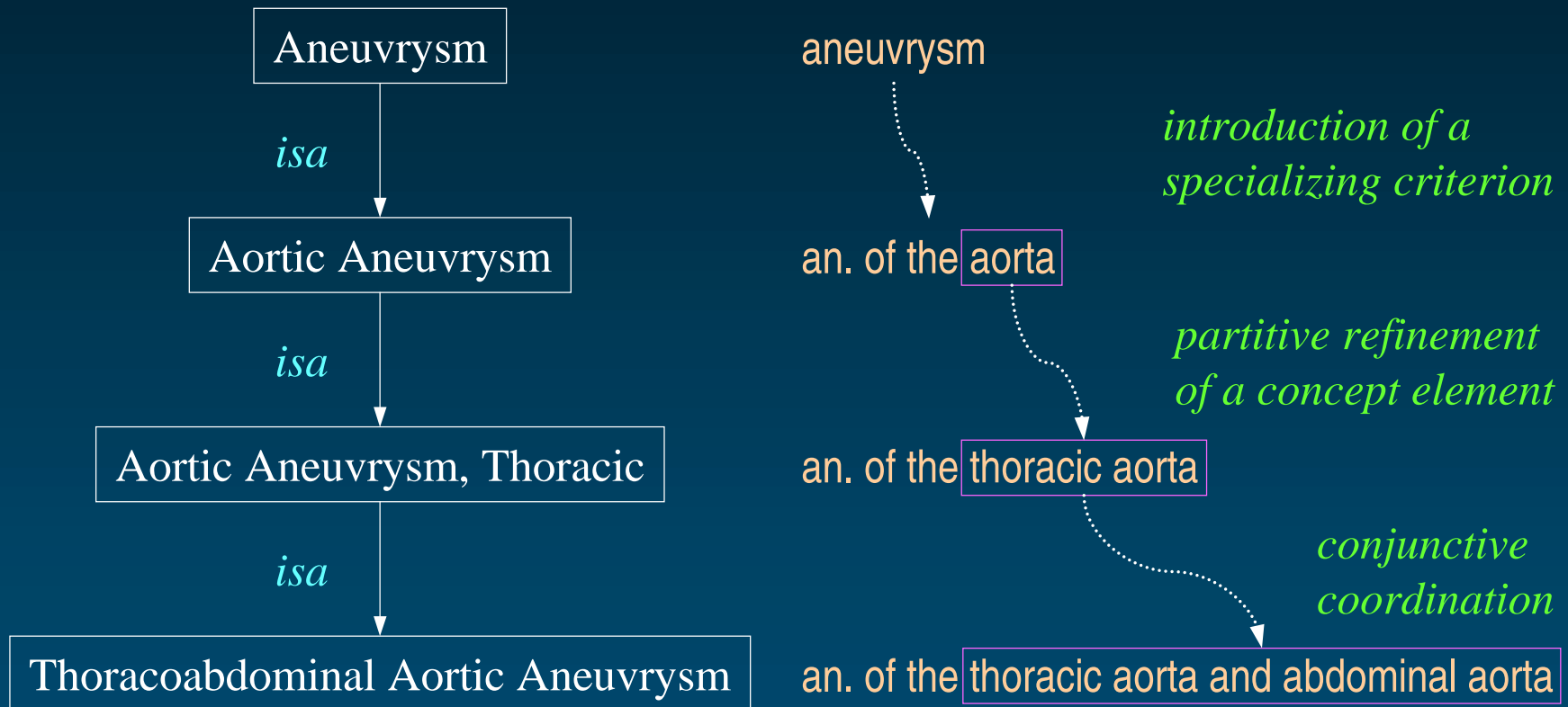
- [Reflexivity]
- Antisymmetry
- Transitivity

## ◆ Inheritance

## ◆ Reasoning



# Principles of subsumption



# Hierarchies in source vocabularies

---

## ◆ Structure

- Single tree
- Polyhierarchical (multiple parents)

## ◆ Relationships

- Usually implicit
- May be other than *isa* or *part of*
  - E.g., Thesaurus relationships

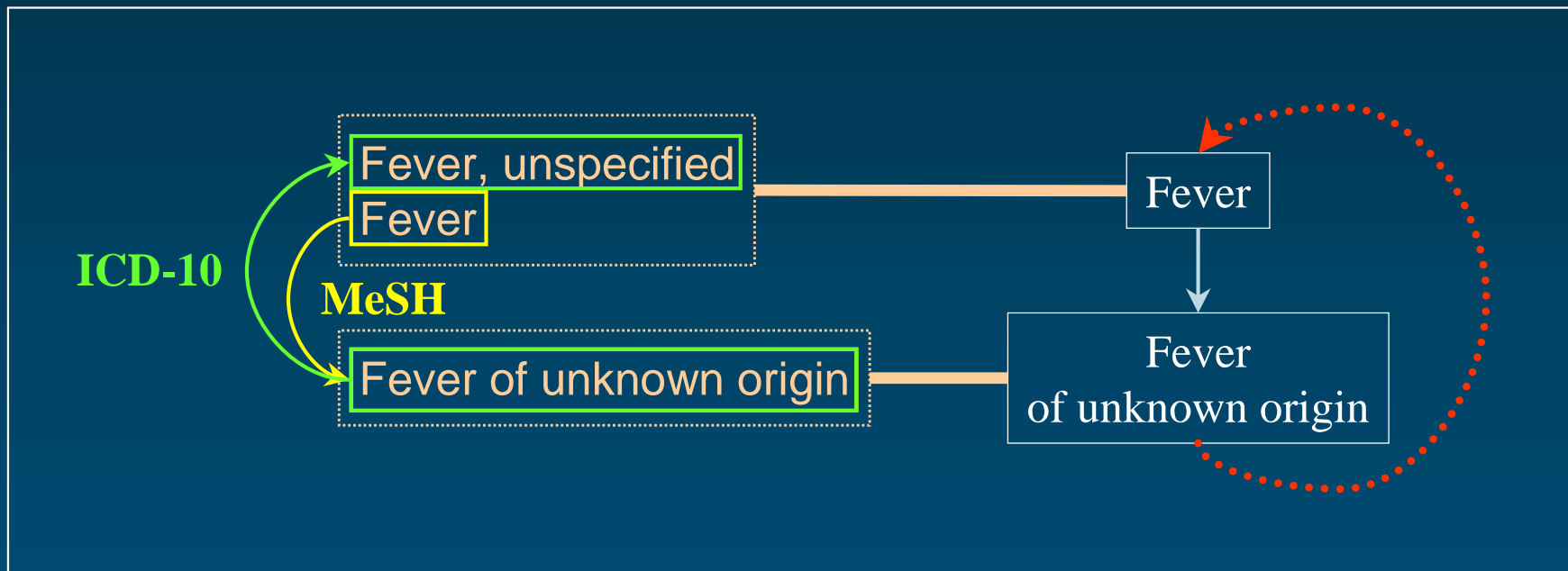
# Hierarchies in source vocabularies

---

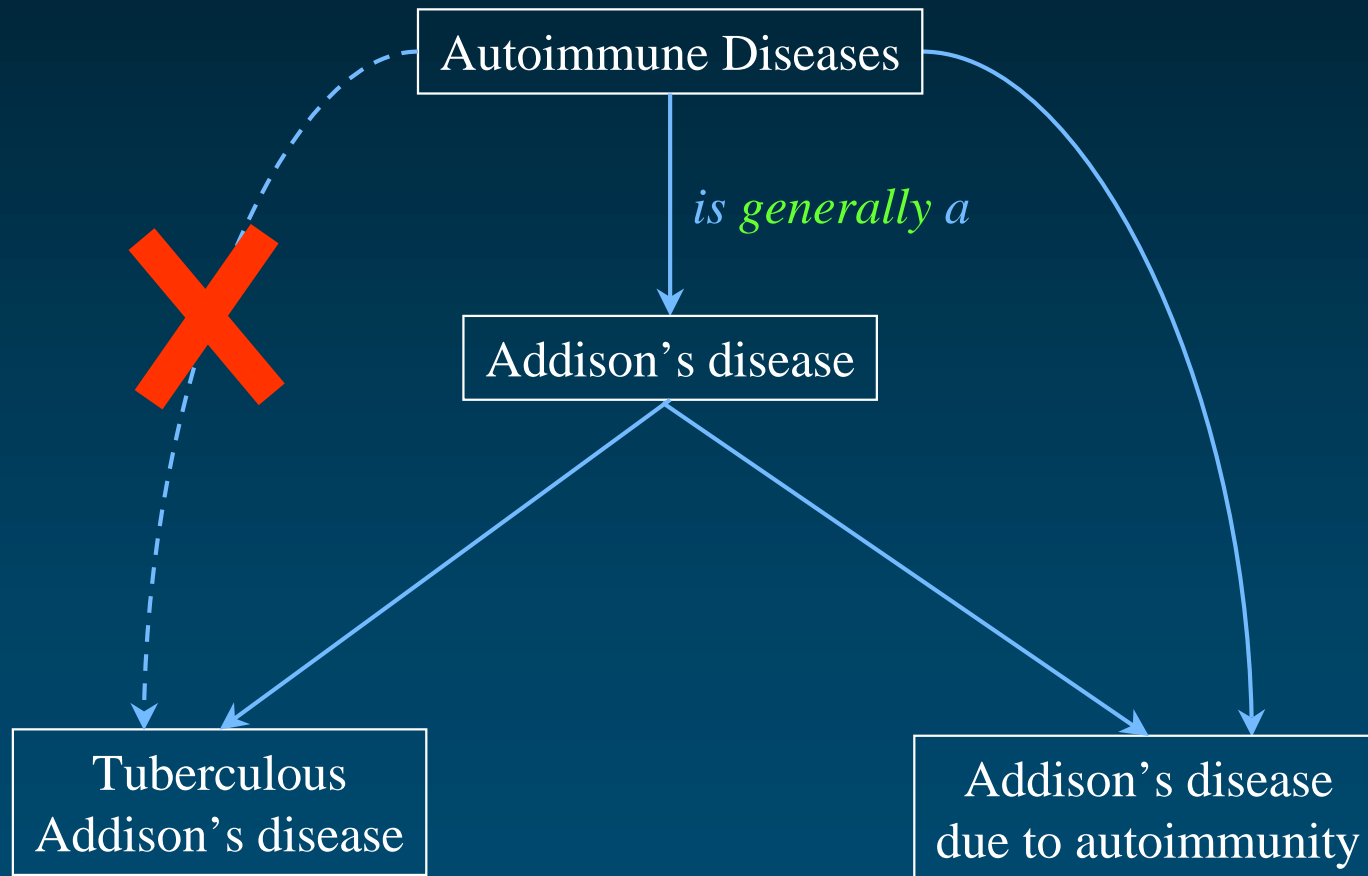
- ◆ Often task-driven  
rather than based on principles
- ◆ Usually suitable for information retrieval
  - Better recall
  - Precision may not be crucial
- ◆ Generally not suitable for reasoning

# Some “isa” relationships are symmetric!

- ◆ Specified and underspecified terms
  - May appear at different levels in a source hierarchy
  - Are clustered into the same concept (same meaning)



# Not all “isa” relationships are transitive!

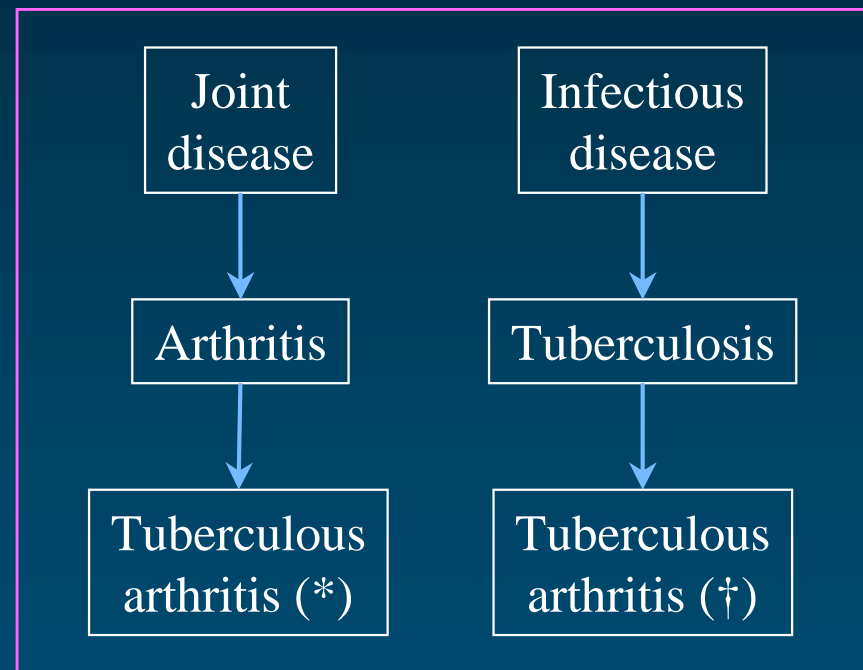
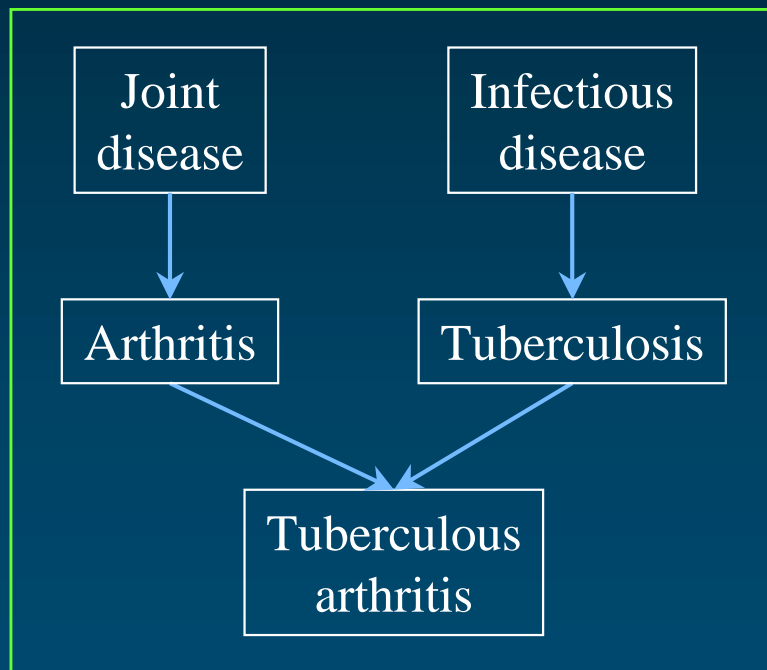


# Mapping relationships (Cross-reference)



# Internal mapping relationships

- ◆ Used to implement multiple inheritance in monoaxial systems



# External mapping relationships

---

- ◆ Created to translate between medical vocabularies
  - SNOMED to ICD-9-CM
  - MeSH Supp. concepts to Descriptors
  - ICD-9-CM to MeSH
- ◆ Weak semantics
  - IR relationships
  - Clinical relationships (“clinically associated with”)

# Alternative approaches

# Concepts

---

- ◆ Post-coordination
  - atomic concepts
  - compositionality
- ◆ Model-oriented terminologies
  - conceptual structures
  - description logics

# Terminologies

---

## ◆ GALEN

- EU project (Rector & al.)
- Strong emphasis on the model
- Limited coverage
- Support for terminology services

## ◆ SNOMED-RT/CT

- Developed by CAP
- Based on SNOMED
- Additional structure (reference terminology)
- Collaboration with CTV3 (formerly Read Codes)



# However...

---

- ◆ Less intuitive
- ◆ Steep learning curve
- ◆ Not widely deployed yet
- ◆ Very expensive to build
- ◆ Less comprehensive

# Conclusions

# Conclusion

---

- ◆ Biomedical domain is vast
  - Broad and deep
  - 800,000 UMLS concepts
- ◆ Difficult to organize consistently
- ◆ Organization of knowledge needs to be principled to support reasoning
- ◆ Evolve
  - From structures supporting IR
  - To structures supporting reasoning



# References

---

Bodenreider O., Bean C. A.

Relationships among knowledge structures: Vocabulary integration within a subject domain.

in Bean, C. A. and Green, R. eds. *Relationships in the organization of knowledge*, Kluwer, 2001, p. 81-98.



Contact: [olivier@nlm.nih.gov](mailto:olivier@nlm.nih.gov)



*Olivier Bodenreider*

Lister Hill National Center  
for Biomedical Communications  
Bethesda, Maryland - USA